



Llengua i societat de la informació: construint una capa de recursos lingüístics lliures pel català 08/01/2004, Jordi Mas i Hernández, jmas@softcatala.org

Sovint em pregunten quins són els problemes per al desenvolupament del català a la societat de la informació. La majoria de la gent espera una resposta que apunti cap a les multinacionals, que algunes, certament, tenen la culpa de diverses de les situacions, però la meua resposta és sempre la mateixa: nosaltres mateixos. L'argument mereix una explicació.

La transició que estem vivint actualment cap a una societat de la informació ha modificat profundament les relacions entre les persones, empreses, i governs: les empreses usen la Xarxa per relacionar-se amb els clients i proveïdors, utilitzen també eines de gestió del coneixement per ser més eficients, els governs milloren la seva presència a Internet i els serveis als ciutadans, a través de la Xarxa, els usuaris usen les eines per millorar les seves relacions personals, etc. Anem de manera imparable cap a una societat altament interconnectada on l'eix fonamental és la informació i on les llengües tenen un paper molt destacat.

Perquè el català, com a llengua, pugui afrontar amb èxit aquest repte, necessitem tot un seguit d'eines de recuperació de la informació i de tecnologies de la llengua. Aquestes eines, abracen des de sistemes capaços de detectar automàticament la llengua d'un document (i lògicament destriar el català correctament), de fer un autoresum del contingut, o de ser capaços d'elaborar respostes automàticament a preguntes sobre textos, fins a correctors ortogràfics, gramaticals, sistemes de traducció automàtica, de reconeixement de textos escanejats, o sistemes de reconeixement de veu; és a dir, totes les eines necessàries perquè el tractament de la informació en la nostra llengua sigui una realitat.

Mentre que el seny mateix ens indica la importància de disposar d'aquestes eines, el que té realment d'important i potser no tan de trivial, és adonar-nos que aquestes eines han de formar part d'un bé comú, de tota la comunitat lingüística catalana, perquè són eines bàsiques en la nova societat i són d'utilitat pública, i com a tals estan fora de la llei de mercat: han de ser de lliure accés.

La compartició del coneixement és el model sobre el qual han treballat sempre els científics, publicant i compartint els resultats de la seva recerca, creant una capa de coneixement comú, model sobre el qual la societat sempre ha innovat. S'ha de crear una infraestructura pública bàsica d'eines lingüístiques que puguem utilitzar els ciutadans, empreses i les institucions per treballar en català en les noves tecnologies.

De la mateixa manera que tenim carreteres públiques, que garanteixen la comunicació entre un punt i un altre, necessitem que totes aquestes eines necessàries per al tractament de la informació en llengua catalana siguin públiques i estiguin a l'abast de tothom, que formi un bé compartit per tothom. Això no vol dir simplement que aquestes eines siguin gratuïtes, sinó que necessitem que el programari i els recursos lingüístics que les formen, siguin lliures, entenent per lliures que siguin, com a mínim, de domini públic juntament amb el seu codi, i com a desitjable, alliberades amb una llicència de programari lliure.

Molta gent encara confon el fet que un recurs sigui gratuït amb el fet que sigui lliure. Per exemple, el CERCATERM del TERMCAT[1], que recull l'adaptació dels neologismes més recents en llengua catalana, és una base de dades de consulta gratuïta, però no és lliure. I no és lliure perquè no ofereix una descàrrega del contingut de la bases de dades ni una llicència flexible amb la qual es pugui treballar. Això, sense dubte, facilitaria molt la tasca de nombrosa gent que treballem en el món de la traducció en diferents àmbits i ens permetria tenir-ne còpies locals que es poden adaptar i millorar a les necessitats de cada moment.

Un dels pocs recursos lliures de lingüística computacional que hi ha avui en dia pel català és el corrector ortogràfic de Softcatalà elaborat per en Joan Moratinos[2]. El fet que sigui lliure ha fet possible que s'hagi pogut incorporar en les versions en català dels programes Open Office, Abiword, Mozilla, o en el sistema operatiu GNU/Linux. Així mateix, també s'ha pogut acoblar a sistemes de correu electrònic tipus correuweb o a blocs com ara PuntBarra, on per exemple, els usuaris poden corregir les notícies abans d'enviar-les. Sens dubte, l'impacte que ha tingut aquest recurs lingüístic lliure pel català ha estat molt important ja que cadascú l'ha pogut adaptar lliurement a les seves necessitats. El fet que existeixi aquest recurs lliure no impedeix que hi hagi empreses amb productes com ara l'Escriu[3], el Word Correct [4], o el Maxigramar[5] que siguin comercials i no lliures.

Un cas paradigmàtic en l'àrea de recursos lingüístics són els sistemes de traducció automàtica. Aquests sistemes són fonamentals per al desenvolupament del català a les noves tecnologies. Avui en dia tenim casos de molts llocs web amb continguts en castellà o anglès que podrien ser traduïts i adaptats si aquestes eines haguessin estat lliures, ja que facilitarien aquesta tasca enormement i reduirien els costos de traducció. És necessari que siguin lliures no només per tots els motius que hem exposat, sinó també perquè és l'única manera de garantir la flexibilitat que la majoria d'usuaris necessita (incloent-hi personalització del lèxic i els formats d'entrada i sortida) així com garantir una continuïtat de la tecnologia en el futur.

Realment el català podria fer un salt molt important en la seva presència a Internet i en la societat de la informació en general si tinguéssim traductors català - anglès i català - castellà amb tecnologia lliure. El Govern català ha invertit molts euros en tecnologia d'aquest tipus i a casa nostra tenim universitats que treballen amb diners públics, en aquesta àrea de recerca, però que mai han produït cap recurs lliure.

El fet de no poder disposar de totes aquestes eines, bé perquè no existeixen o bé perquè estan en mans d'universitats que no hi donen accés o empreses privades, representa establir una barrera d'accés per al català a la societat de la informació i dificultats afegides als que treballen en altres llengües perquè vulguin apostar pel català.

Aquestes eines s'han de desenvolupar sense rebre diners i sempre de manera altruista. La resposta és no, necessàriament. Una universitat o empresa pot rebre suport d'un programa europeu o del Govern per desenvolupar aquestes eines. El que és molt important és que una vegada fet el projecte sigui un bé comú per a tothom. De fet, per al català algunes eines ja existeixen i han estat desenvolupades amb diners públics però actualment es troben tancades.

L'existència d'aquesta capa comuna de béns per al català en forma de tecnologies per a la llengua deixa espai per a les empreses perquè puguin desenvolupar productes més

especialitzats, personalitzats, i oferir-hi suport. Així mateix, també fomentem la innovació proporcionant una infraestructura bàsica sobre la qual puguem treballar-hi, disminuint les barreres d'entrada de noves empreses que vulguin treballar amb tecnologies en català.

Hem d'evitar que les tecnologies de la llengua estiguin exclusivament en mans del sector privat i hem d'assegurar-ne la disponibilitat com a bé públic evitant així l'establiment, de manera indirecta, d'un impost encobert (en forma de llicències de programari) per treballar en la nostra llengua en la societat de la informació, un fet que avui en dia ja s'està produint.

Els diferents governs dels països catalans, les institucions catalanes, les universitats (especialment els departaments de lingüística computacional) tenen una gran responsabilitat a assegurar que aquesta capa de recursos lliures per al català sigui una realitat, alliberant la recerca bàsica que ja han fet amb diners públics i tenint aquest objectiu sempre present en futurs projectes.

Si no reeixim en la creació d'aquest conjunt d'eines lliures que permetin treballar, a tothom, en català, s'haurà fracassat a portar al nou model de societat de la informació el català .

Enllaços

- [1] <http://www.termcat.net/neoloteca/>
- [2] <http://www.softcatala.org/projectes/myspell/>
- [3] <http://www.inedit.com/>
- [4] <http://www.market-way.net/wordcorrect/>
- [5] <http://www.maxigramar.com/>

Bibliografia recomanada

La lengua que era un tesoro, Jose Antonio Millan
<http://jamillan.com/tesoro.htm>

La importància de la propietat intel·lectual per a l'avanç del català a les noves tecnologies,
Jordi Mas <http://www.softcatala.org/jornades/ponencies/ponencies/002-softcatala.pdf>

Nota: Aquest article va ser publicat en versió reduïda el 15 d'octubre al Diari Avui, així com a la revista Llengua i Ús del mes de desembre.

Aquest document és (c) 2004 Jordi Mas i Hernández. Es permet el seu ús i distribució en qualsevol mitjà sempre que no sigui modificat i s'inclogui aquesta nota.

[[Inici](#) | [Qui som](#) | [Projectes](#) | [Fòrums](#) | [Escriu-nos](#) | [Botiga](#) | [GNU/Linux](#) | [Mac](#) | [Traductor](#) | [Llistes](#) | [Enllaços](#) | [Ajuda](#)]